AD_______________

Award Number: W81XWH-07-1-0710

TITLE: Hormonal Involvement in Breast Cancer Gene Amplification

PRINCIPAL INVESTIGATOR:
Susan A. Gerbi, Ph.D. (P .I.), Alexander Brodsky, Ph.D . (co-P.I.), Ben Raphael, Ph.D. (co-P.I.)

CONTRACTING ORGANIZATION:
Brown University

Providence, RI 02912

REPORT DATE: October 2010

TYPE OF REPORT: Final

PREPARED FOR:
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

XX Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 01-10-2010	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* 24 Sep 2007 - 23 Sep 2010

4. TITLE AND SUBTITLE Hormonal Involvement in Breast Cancer Gene Amplification	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER W81XWH-07-1-0710
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Susan A. Gerbi (P.I.) Alexander Brodsky (co-P.I.) Ben Raphael (co-P.I.)	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Brown University; Providence, RI 02912	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, MD 21702-5012	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT (ca. 200 words)

Our specific aims were to map origins of replication in the breast cancer genome and compare these sites with genomic positions that bind the estrogen receptor (ER) and genomic regions of DNA amplification. Correlations would support our hypothesis that ER adjacent to replication origins may interact with the replication machinery to drive DNA amplification, hallmark of many cancers. We began our experiments with the well-studied breast cancer cell line MCF7. We refined the lambda exonuclease method to obtain nascent DNA strands (up to 20 times enriched) and demonstrated that the myc origin maps to the same position in HeLa and MCF7 cells. We size fractionated nascent DNA (500-1500 nt) for next generation sequencing which gives better resolution and sensitivity than DNA microarrays. The results of Illumina sequencing revealed that our nascent DNA sample from the whole genome included many of the origins previously reported for 1% of the human genome. Development of this technology serve as the basis for our IDEA Extension grant which was funded to extend the work begun in this parent grant.

15. SUBJECT TERMS
estrogen receptor, DNA amplification, replication origins

16. SECURITY CLASSIFICATION OF: U			**17. LIMITATION OF ABSTRACT** UU	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSO** USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U		40	**19b. TELEPHONE NUMBER** *(include are code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

FINAL PROGRESS REPORT

INTRODUCTION

Recent data support the hypothesis that DNA amplification plays a role in establishing the malignant cell phenotype in cancer (Nikolsky et al., 2008). However, the basic mechanism underlying DNA amplification has not yet been elucidated, though it may be more common than originally thought (Gomez 2008; Gomez and Antequera 2008). There appears to be a link between the steroid hormone estrogen and many forms of breast cancer, but the detailed mechanism is unknown. Estrogen can turn on gene expression and thus activate the production of the proteins encoded by these genes. Our recent results in a model system indicated that a steroid hormone can induce gene amplification in which re-replication creates extra copies of the gene. This in turn will also increase production of the protein encoded by the amplified gene. Hormonal induction of gene amplification is a new paradigm for how hormones work, and we wish to see if it applies to breast cancer. <u>We wish to examine if a correlation exists between estrogen receptor (ER) binding at novel sites in the breast cancer genome and juxtaposition with replication origins that escape normal cellular controls and re-replicate, leading to DNA amplification.</u> The recent observations that estrogen induces cell proliferation by retention of MCM proteins in the nucleus and by induction of the loading factor Cdt1 (Pan et al., 2006) support our hypothesis, especially since increases in MCM proteins and Cdt1 have been shown to induce DNA amplification in yeast (Gopalakrishnan et al., 2001; Nguyen et al., 2001; Green et al., 2006) and increased Cdt1 results in re-replication in human cells (Dorn et al., 2008). The N-terminus of Cdt1 is important for re-replication, perhaps through interactions with PCNA and/or cyclin (Teer and Dutta, 2008). Cdt1 and its inhibitor geminin are deregulated in human tumors (Petropoulou et al., 2008). Moreover, stalled replication forks and DNA re-replication lead to DNA breakage and rearrangements (Green and Li, 2005; Raveendranathan et al., 2006; Zhu and Dutta, 2006; Dutta, 2007; Hook et al. 2007) which is a hallmark of cancer. Our research may provide a new paradigm for hormonal induction of breast cancer via gene amplification, leading to new methods of diagnosis and treatment.

BODY

In the research supported by this grant, we proposed to map estrogen receptor binding sites, origins of replication and regions of DNA amplification in surgically derived breast cancer tissue (see Appendix 1: DOD meeting abstract). We report our progress on these three specific aims. Also, we report on relevant recent publications that support our working model. The P.I. (Susan Gerbi) and two co-P.I.s (Alex Brodsky and Ben Raphael) meet together with their lab personnel roughly once a month to review past results and design future experimental strategies.

The text that follows is the revised final report for this grant, organized according the Statement of Work that was listed in the approved grant application which was organized according to the three Specific Aims. Figures and tables are located at the end of the document.

Statement of Work Item 1: Map ER binding in the human genome

Months 1-6: Work out the methodology (Gerbi and Brodsky labs)

Chromatin from Breast Cancer Tissue - this is the starting material for each of the three specific aims. In our previous progress report we reported the difficulty in obtaining breast cancer tissue samples, as recent changes in clinical protocols now result in the vast majority of patients receiving chemotherapy prior to breast cancer surgery. We cannot use tissue derived from patients with neoadjuvent therapy, and therefore, our potential supply of material was drastically reduced. As reported previously, we expanded our network of clinical collaborators to include other surgeons and pathologists at both R.I. Hospital and Women and Infants Hospital. With this increased outreach, we have now obtained some surgically derived tissue samples and are hoping for more samples (see **Table 1**). To work out the methodology for chromatin isolation from breast cancer tissue, we used material from the R.I. Hospital breast cancer tumor bank that did not meet our criteria above, but was available for these pilot experiments. We determined that tissue that was freshly obtained and frozen from a current surgery was equivalent to tissue that had been stored frozen for a period of time. For chromatin immunoprecipitation (ChIP) procedures, the tissue is subjected to formaldehyde fixation, homogenized and then sheared by sonication. We found that the breast cancer tissue (portions of 1.0-1.5 cm tumor specimens) was very fibrous and hard to break open by standard homogenization, resulting in low yields of chromatin. We purchased a microhomogeneizer (the same model used by Dr. Peggy Farnham for her breast cancer chromatin studies) to use for breast cancer tissue disruption. Our initial results were promising. We were able to prepare sonicated chromatin from cell lines and tissues using the microhomogenizer averaging less than 500 bp in size (**Figure 1**) and that works well for ChIP.

Months 7-20: Carry out ER ChIP-chip experiments

(Brodsky, Gerbi and Raphael labs)

ChIP of ER binding sites in the breast cancer genome - these data had already been obtained by co-P.I. Alex Brodsky for MCF7 cultured breast cancer cells (Carroll et al., 2005 and 2006), and we planned to use them as a reference source as we developed the ChIP methodology for breast cancer tissue. In the grant application we proposed to do ChIP-chip (DNA microarrays of chromatin immunoprecipitated samples). However, that method has now been superceded by ChIP-Seq (DNA sequencing of chromatin immunoprecipitated DNA samples). This method has greater sensitivity, in addition to its better resolution than ChIP-chip. ChIP-Seq revealed 10-13,000 ER-alpha binding sites in the genome of MCF7 breast cancer cells (Lin et al., 2007; Fullwood et al., 2009; Welboren et al., 2009; Hurtado et al., (2011). Because of the difficulty in obtaining tumor sample specimens (see preceding section) and the fact that none of the samples we received were FISH positive for HER2 gene amplification (**Table 1**), we decided to carry out our experiments MCF7 cultured breast cancer cells. An added bonus of this cell culture model system is that the ER-alpha binding sites have already been mapped by several groups by the more advanced method of ChIP-Seq (see above), obviating the need for us to map these sites.

Months 21-24: confirm ER binding site candidates (Gerbi lab)

This was no longer necessary since other groups have already mapped ER-alpha binding sites in the MCF7 breast cancer genome (see above). The decreased experimentation needed for statement of work item 1 allowed us to devote more time to develop methodology for statement of work item 2 (see below).

Statement of Work Item 2: Map binding sites for the Origin Recognition Complex (ORC) across the human genome (Brodsky, Gerbi and Raphael labs)

Months 7-20: Map replication origins in the breast cancer genome

As stated above, because of the paucity of breast cancer tissue material, we decided to do our initial experiments to map replication origins on the well studied MCF7 cell line where the ER binding sites are already mapped. This also has the advantage of sample homogeneity which would be a concern for tissue from breast cancer tumors. Our plan was to analyze DNA from unsynchronized cells in order to capture all origins regardless of when during S phase they are activated. We reported in previous progress reports that we obtained polyclonal antibodies for human ORC2 and Cdt1 and a monoclonal antibody against human ORC6 from Aloys Schepers, antibody against human ORC1 from Mel DePamphilis and a mammalian expression clone for FLAG-tagged human ORC1 from Dr. Kohji Noguchi (a former post-doc with Dr. Mel DePamphilis). The Brodsky lab checked these antibodies by Western blots, but there was high background with multiple bands. Moreover, two tries of ChIP with ORC antibody were unsuccessful. Discussions that P.I. Susan Gerbi had with Drs. Aloys Schepers and Michael Leffak at the Cold Spring Harbor DNA Replication Meeting revealed that ChIP on mammalian cells with ORC2 antibodies has a high background. Instead, the Brodsky lab considered cloning ORC1 into a FLAG-tag vector for transfection into MCF7 breast cancer cells, reasoning that ChIP with a FLAG antibody should give better results. However, we decided instead to pursue the more promising approach described below.

The goal is to map DNA replication origins in the breast cancer genome. A problem with the approach of mapping ORC binding sites is that ORC also binds to silent origins that are not used, so we would not know which origins are active in the breast cancer cells. Moreover, as described above, there are problems with the antibodies against ORC. We decided that superior results would be obtained with a more successful approach to isolate small nascent DNA to map replication origins directly by sequencing the nascent strands, rather than using ChIP-chip or ChIP-seq to map ORC binding sites. The short nascent strand sequencing approach allows us to identify by this direct method all origins that are active in the breast cancer genome. Nascent strands have been used to map replication origins for a limited portion (1% ENCODE project) of the human genome (Lucas et al., 2007; Cadoret et al., 2008) and has given more reliable results than BrdU labeling of non-lambda exonuclease treated DNA (Birney et al., 2007; Karnani et al., 2007) where results from the latter do not agree with results of mapping replication bubbles trapped in agarose (Mesner et al., 2010).

In order to identify origins of DNA replication throughout the genome of MCF7 cells, nascent DNA was prepared according to our previous protocol (Gerbi and

Bielinsky, 1997).As summarized in **Figure 2**, genomic DNA was prepared from mid-log phase cells using DNAzol (Invitrogen, Calsbad, CA) and resuspended in TE. Replicative Intermediate (RI) DNA was enriched by passing the genomic DNA over a column of BND-cellulose. The ends of the RI DNA were phosphorylated using T4 Polynucleotide Kinase (New England Biolabs, Ipswich, MA). Next the DNA was digested with lambda exonuclease to enrich nascent strands which are resistant to lambda-exonuclease digestion due to the presence of an RNA primer at their 5' end. Finally, the nascent strands were size fractionated (500 – 1500 bp) on low melting point agarose to eliminate background from Okazaki fragments which occur throughout the genome. Enrichment of nascent strands was confirmed by real-time PCR assaying for enrichment of the c-myc origin of replication (Tao et al., 2000). Primers were designed to be specific to locus 11 (the c-myc origin) and a non-origin sequence about 6 kb upstream at locus 1. The origin mapping experiments are being carried out by Dr. Michael Foulk, a talented postdoc in the Gerbi lab. Due to the fact that the c-myc origin of replication was discovered in HeLa cells, we first compared the enrichment of nascent strands between HeLa cells and MCF7 cells in order to confirm that the replication origin maps to the same position at the c-myc locus in MCF7 cells. Our results confirmed that this was indeed the case: In HeLa cells, the c-myc origin of replication was enriched about 12 fold, while in the MCF7 cells, it was enriched about 11 fold when we used the DNA nascent strand isolation protocol above, suggesting an origin of replication exists at the same locus in MCF7 cells (**Figure 2**). These experiments also demonstrated the feasibility of isolating nascent strands from MCF7 cells for further analysis. Subsequently, we were able to reliably obtain ~ 100-150 ng nascent strands from 100 ug starting genomic DNA from asynchronous MCF7 cells. Real-time PCR showed that the preps were enriched for the c-myc origin between 11.0 and 19.6 fold (**Figure 2**).

Several preparations of nascent DNA from MCF7 cells were pooled and submitted for next generation sequencing. Ben Raphael aligned the resulting Illumina reads to the human genome using MAQ resulting in 5.6 million mapped reads. We counted the number of reads that align to genomic intervals defined by 283 replication origins identified in 1% of the human genome (Cadoret et al., 2008; ENCODE project). For each of these 283 intervals, we compared the read count of the interval to the expected read count under a uniform distribution of reads to intervals. We found that 78 of the 283 origins were enriched ($P<10^{-3}$) for nascent strands (**Figure 3**; **Table 2**). These initial results were encouraging, and suggest that our nascent DNA preparation is enriching for replication origins. However, our sequencing coverage was not high enough to robustly detect new replication origins and there was some contaminating bacterial DNA.

However, these experiments proved difficult to reproduce. We determined that the cause of this variability was in the poor quality of the preparation of lambda exonuclease we were using (our previous source from Invitrogen had been discontinued so we had switched to enzyme from New England BioLabs). By discussion with Drs. Mechali and Prioleau whose labs are in France, P.I. Susan Gerbi learned that the company Fermentas could prepare high quality lambda exonuclease by special order.

Therefore, we contracted with Fermentas to obtain a high quality, high concentration preparation of lambda exonuclease. The original nascent strand protocol was modified so that the phosphorylated RI DNA was digested with 240 units of lambda exonuclease (versus 15 units previously) overnight following a protocol developed by Cadoret et al. (2008) for mapping replication origins in the ENCODE subset of the human genome. Using this protocol we achieved about 20 fold enrichment of the c-myc origin in MCF 7 cells, which is excellent. We then, however, noticed some variability in nascent strand enrichment from preparation to preparation. We did controls that revealed that the pH optimum for the Fermentas recombinant lambda exonuclease is broader than for the previous Invitrogen purified enzyme. At the previously used higher pH of 8.8 we found that there was degradation of RNA; this would compromise the integrity of the RNA primers on the nascent DNA and render the nascent DNA susceptible to lambda exonuclease digestion (**Figure 4**). We found that the Fermentas preparation of lambda exonuclease is still active at pH 8.0 and that RNA degradation does not occur at that pH (**Figure 4**). Additionally, we found that heating the samples resulted in degrading the RNA so we have modified the original protocol to eliminate heating steps where possible (**Figure 4**).

For the Illumina sequencing, we isolated about 50-150 ng nascent strands from 100 ug starting genomic DNA. Several nascent strand preparations were pooled (about 500 ng total) and subsequently amplified at the Yale University sequencing facility for Illumina sequencing. However, it required a lot of effort to obtain this amount of nascent DNA. Also, PCR artifacts can be introduced during the amplification step. To overcome these problems, the DOD IDEA Extension grant will allow us to try Helicos rather than Illumina sequencing. The first report of Helicos sequencing appeared just a year ago (Harris et al. 2008) and holds much promise (Gupta 2008). This true single molecule sequencing (tSMS) approach omits the necessity for DNA amplification, significantly reducing the amount of nascent DNA starting material required, about 10 ng. In comparison, the Illumina platform for complete genome coverage requires 500-1000 ng nascent DNA. Moreover, since the nascent DNA preparation enriches for the single stranded leading strands near an origin of replication the sequences should map to opposite strands on either side of the origin. This data will provide a signature for authentic origins of replication reducing the potential for calling false positives. There are only a few Helicos machines in operation world-wide. We have been given access on a fee-for-service basis to the Helicos sequencing machine at the Dana Farber Cancer Institute where one of us (Alex Brodsky) was previously a postdoc prior to joining the faculty at Brown University. We intend to also sequence the nascent DNA using Illumina and compare the results between the two platforms. We anticipate that there will be ~25,000-30,000 replication origins in the human genome. The replication origins, once mapped, will be compared to estrogen receptor binding sites (the ENCODE data showed a correlation for c-JUN and c-FOS as potential regulators of origins; Cadoret et al., 2008) and to regions of DNA amplification in breast cancer cells.

To sum up, we have accomplished much more for Statement of Work Item 2 than originally presented in the grant proposal. Instead of using ChIP-chip or ChIP-Seq to map ORC binding sites, we have refined the methodology to allow direct mapping of all active origins in the breast cancer genome by isolation of short nascent strands of DNA and sequencing them. Our pilot run on the Illumina platform was successful. Based on this progress, we are grateful to have received a DOD IDEA Extension award.

Statement of Work Item 3: Mapping DNA amplification sites in the breast cancer genome

One of us (Ben Raphael, co-P.I.) and others have identified chromosomal changes in the genome of MCF7 breast cancer cells (Volik et al. 2003 and 2006; Raphael et al., 2008; Hampton et al., 2008), including sites of DNA amplification. Therefore, once we have mapped the replication origins in MCF7 cells (Statement of Work Item 2), we can directly compare their locations with the already mapped locations of ER binding sites (Statement of Work Item 1) and DNA amplification (Statement of Work Item 3), as per our specific aims for this grant. If we extend the study to breast cancer tissue, array comparative genome hybridization (aCGH) will be used to determine the sites of DNA amplification in this tissue.

Months 1-6: Refine methods to analyze DNA amplification (Raphael lab)

DNA double strand breaks have been shown to play a role in DNA amplification. As stated in our previous progress report, co-P.I. Ben Raphael developed a novel method called Neighborhood Breakpoint Correlation (NBC) to identify correlated rearrangement breakpoints from CGH data (*BMC Bioinformatics*, in revision). Unlike previous methods for aCGH analysis that focus on finding common genomic intervals of amplification or deletion that might harbor oncogenes or tumor suppressor genes, respectively, NBC focuses on the precise localization of the boundaries (breakpoints) of these intervals. We hypothesize that pairs of such highly conserved breakpoints might indicate fusion genes or other common rearrangements. The algorithm employs a statistical model derived from the binomial distribution to assess the statistical significance of breakpoints that shared by multiple patients. The algorithm also identifies genes or pairs of genes that each contains one or more breakpoints in a statistically significant number of patients.

In preliminary analysis, Ben Raphael examined a collection of 36 primary prostate tumors for breakpoints in the well-known TMPRSS2-ERG fusion gene (Tomlins et al. 2005). He applied NBC to identify changes in copy number (breakpoints) in each patient and then identified common breakpoints that appear in a statistically significant number of patients. Specifically, he identified 12 statistically significant rearrangements, one of which is the TMPRSS2-ERG fusion gene. It is detected in 5 patients with a p-value of 2. 7x10^10 (**Figure 5**). In a larger analysis, he examined a collection of data from 233 patients with glioblastoma, including 227 primary tumor samples and 107 matched blood samples from The Cancer Genome Atlas. He predicted 93 statistically significant rearrangements that are further classified as gene truncations, germline structural variants, and fusion genes. The power of his method to detect correlated breakpoints increases with larger sets of patients. We will apply these methods to the aCGH data that will be generated in the present research project on DNA amplification in breast cancer cells. This will allow us to uncover additional candidate fusion genes or regulatory fusions, particularly fusions near ER binding sites.

We now report here some new data derived by computational analysis by co-P.I. Ben Raphael. He combined aCGH data and ESP data from estrogen receptor (ER) binding data in MCF7 breast cancer cells determined initially by co-P.I. Alex Brodsky using chromatin immunoprecipitation (ChIP-chip) (Carroll et al. 2006). The original ChIP-chip study (Carroll et al., 2006) interpreted the data in the context of the reference human genome, even though it is well known that MCF7 exhibits extensive genomic aberrations including copy number changes and structural rearrangements. We examined how knowledge of these genomic changes affects the interpretation of the ChIP-chip data. Using scan statistics (Glaz et al., 2001), we identified regions of the reference genome that contained significantly few or significantly many ER sites. We found 38 gaps, defined as sequenced genomic regions > 6.9 Mb with no ER binding sites. Under the null hypothesis of ER sites distributed uniformly on the genome, the probability of finding one such gap is < 10^{-4}. The copy numbers of the probes in these gaps determined using Agilent 44K array CGH had mean $\log_2$-ratio of -0.28, significantly below the mean value of -0.05 over all probes (p-value by T-test < 10^{-100}) and implying that the gaps in ER binding sites are a result of deletions in the MCF-7 genome.

We next identified unusual clusters of ER binding sites in the ChIP-chip data, where a cluster was defined as 12 or more ER sites in a 1 Mb region of the genome. Under the null hypothesis, the probability of such a cluster is < 6×10^{-5}, but we identified 11 clusters in the data. The mean $\log_2$-ratio in these segments is 1.3, suggesting that clusters of ER binding sites are preferentially found in amplified regions. This could be due to the fact that the ChIP-chip assay has higher sensitivity in amplified regions, or due to the fact that the statistical model used to call binding sites is imprecise in amplified regions. To assess whether there might be amplification of the regions harboring ER clusters, we examined BAC array CGH data on 51 breast cancer cell lines from (Neve et al., 2006). We found that one of the 11 ER clusters is preferentially amplified in ER-positive cell lines and preferentially deleted in ER-negative breast cancer cell lines (**Figure 6**). Moreover, we found differential expression of one of the four genes in this region using Oncomine (Rhodes et al., 2007). TLE3 is significantly under-expressed in ER-negative breast tumors compared to ER-positive breast tumors (Sotiriou et al., 2003; Minn et al., 2005; Wang et al., 2005; Hess et al., 2006) (p-values 10^{-12}, 2×10^{-11}, 2×10^{-10}, 8×10^{-9} based on t-test). This result is consistent with the amplification of this region as determined by array CGH, and also suggests estrogen dependent regulation of this gene.

We also looked for structural rearrangements that might yield regulatory fusions between bound ER sites in the ChIP-chip data and genes identified as differentially regulated in response to estrogen. Using ESP data to identify rearrangements (Volik et al., 2006; Raphael et al., 2008), we found 27 examples of such candidate fusions ($P < 2 \times 10^{-11}$ by a permutation test.). Moreover, in 11/27 cases, the genes involved in these putative fusions have no bound ER site within 100 kb of the transcription start site. Several of these genes are implicated in breast cancer including BRCC3 (a subunit in the BRCA1/2 containing protein complex), PTK6, and STK6.

Months 21-24: Whole genome SNP arrays (Brodsky and Gerbi labs)

Having shifted from tumor specimens to MCF7 breast cancer cells, SNP arrays were no longer needed.

To sum up, we have done more work in the Statement of Work Item 3 than originally described in the grant application, having developed many tools for computational analysis.

Concluding remarks – The recent finding that the transcription factor c-Myc interacts with the pre-replication complex to control DNA replication (Dominguez-Sola et al., 2007; Lebofsky and Walter, 2007) and that the androgen receptor interacts with MCM7 of the pre-replication complex (Shi, 2008) provides precedence for our hypothesis that the ligand-bound estrogen receptor may play a direct role in regulating replication origins beyond its traditional role as a transcription factor. We are grateful for the DOD funding that allowed us to initiate experiments to test our hypothesis and look forward eagerly to results from the DOD funded IDEA Extension award.

KEY RESEARCH ACCOMPLISHMENTS

- Further refinement in the method to isolate nascent (newly replicated) DNA, lowering the pH to prevent RNA degradation during lambda exonuclease digestion, thereby reducing the prep-to-prep variability in origin enrichment.

- PCR mapping of the myc replication origin showed that it is located in the same position in HeLa and MCF7 cells.

- A trial run of Illumina sequencing of nascent strands included many of the replication origins previously reported for 1% of the human genome

- Improvement of the methodology for analysis of aCGH data to identify common aberrations and common breakpoints.

- Computational analysis suggested that clusters of ER binding sites are preferentially found in amplified regions.

REPORTABLE OUTCOMES

- Method to isolate nascent (newly replicated) DNA

- Preliminary data mapping replication origins in the breast cancer genome

- Methodology for analysis of aCGH data to identify common aberrations and common breakpoints.

- Computational results suggesting that clusters of ER binding sites are preferentially found in amplified regions.

We anticipate writing a paper for publication in a high profile journal describing the map of replication origins in the entire human genome (from breast cancer cells) once our data are complete.

Based on our successful upward trajectory with these experiments, we have received a DOD Idea Expansion award. This award would allow us to complete and expand our promising experiments.

CONCLUSION

Recent publications cited in this progress report support our hypothesis that the estrogen receptor may interact with the replication machinery and promote DNA amplification in breast cancer cells. We have improved the experimental protocol from what was initially approved in this grant. Instead of identifying origins by ORC ChIP, we are isolating size-fractionated nascent strands to use them for next generation sequencing. Our results will be the first to map replication origins on the entire human genome. The data will be compared to map positions of ER binding sites in the genome and regions of DNA amplification. A positive correlation will directly support our hypothesis and will provide a new way of thinking about the role of steroid hormones in cancer. The results will begin to elucidate the mechanism of induction of DNA amplification and could provide a platform for new methods of diagnosis and treatment of breast cancer.

Personnel paid from this grant:
(nb – these are personnel in the three different groups of PI and co-PIs; some of the lab personnel only worked briefly on this project)

PI: Susan A. Gerbi

Co-P.I.s: Alexander Brodsky, Benjamin Raphael

Postdoctorals: Michael Foulk, Yutaka Yamamoto

Graduate Students: Crystal Kahn, Anna Ritz

Research Assistants: Jacob Bliss, Megan Frayne, Mark Grabiner, Sara Hillenmyer, Ingrid Mercer, Shellee Morehead, Hannah Sanford, Heidi Smith

Undergraduate Dishwashers: Carolyn Crisp, Sydney Ember, Emily Hartman, Theeradej Thaweerattanasinp

REFERENCES

Ben-Dor A., Lipson D., Tsalenko A., Re imers M., Baumbusch L.O., B arrett M.T., Weinstein J.N., BΦrresen-Dale A., Yakhini Z. (2007). Framework for identifying common aberrations in DNA copy number data. Inte rnational Conference on Research in Computational Molecular Biology (RECOMB) 20078: 122-136.

Birney E., Stamatoyannopoulos J.A., Dutta A., et al. (2007). Identif ication and analys is of functional element s in 1% of the hum an genom e by the ENCODE pilot project. Nature 447: 799-816.

Cadoret J.C., Meisch F., Hassan-Zadeh V. , Luyten I., Guillet C., Duret L., Quesneville H. and Prioleau M.N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Nat. Acad. Sci. 105, 15837-15842.

Carroll J.S., Liu X.S., Brodsky A.S., Li W., Meyer C.A., Szary A.J., Eeckhoute J., Shao, W., Hestermann E.V.,Geistlinger T.R., Fox E.A., Silver P.A. and Brown M. (2005). Chromosome-wide mapping of estrogen recept or binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122: 33-43.

Carroll J.S., Meyer C.A., Song J., Li W., Ge istlinger T.R., Eeckhoute J., Bro dsky A.S., Keeton E. K., Fertuck K.C., Hall G.F., Wang Q., Bekiranov S., Sementchenko V., Fox E.A., Silver P.A., Gingeras T.R., Liu X.S. and Brown M. (2006) . Genome-wide analysis of estrogen receptor binding sites. Nature Genetics 38: 1289-1297.

Dominguez-Sola D., Ying C.Y., Grandori C., Ruggiero L., Chen B., Li M., Galloway D.A., Gu W., Gautier J. and Dalla-F avera R. (2007). Non-transcripti onal control of DNA replication by c-Myc. Nature 448: 445-451.

Dorn E.S., Chastain P .D. 2nd, Hall J.R. and Cook J.G. (2008) . Analysis of re-replication from redregulated origin licens ing by DNA fiber spreading. Nucleic Acids Res. Nov. 14 (Epub ahead of print).

Dutta A. (2007). Chaotic license for genetic instability and cancer. Nature Genet. 39: 10-11.

Fullwood MJ, Liu MH et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462: 58-64.

Gerbi S.A. and Bielinsky A-K. (1997). Replication initiation point mapping. Methods. 13, 271-280.

Glaz J, Naus JI and Wallenstein S. (2001). *Scan statistics*. New York: Springer.

Gomez M. (2008). Controlled r ereplication at DNA replic ation orig ins. Cell Cycle 7: 1313-1314.

Gomez M. and Antequera F. (2008). Overrep lication of short DNA regions during S phase in human cells. Genes Dev. 22: 375-385.

Gopalakrishnan V., Simancek P., Houchens C., Snaith H.A., Frattini M.G., Sazer S. and Kelly T.J. (2001). Redundant control of rer eplication in fission y east. Proc. Nat. Acad. Sci 98: 13114-13119.

Green B.M. and Li J.J. (2005). Loss of re replication control in Sac charomyces cerevisiae results in extensive DNA damage. Mol. Biol. Cell 16: 421-432.

Green B.M., Morreale R.J., Ozaydin B., Derisi J.L. and Li J.J. (2006). Genome-wide mapping of DNA synthesis in Saccharomyces cerev isiae reveals that m echanisms preventing reinitiation of DNA replicati on are not redundant. Mol. Biol. Cell 17: 2401-2414.

Gupta PK (2008). Single molec ule DNA se quencing technologies for future genomics research. *Trends Biotech* **26**: 602-611.

Hampton O.A., Den Hollander P., Miller C.A. et al. (2008). A sequence- level map of chromosomal breakpoints in the MCF-7 breast cancer cell lin e y ields ins ights into the evolution of a cancer genome. Genome Res. (Epub. Dec. 3, 2008).

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Bras lavsky I, Causey M, Colonell J, Dimeo J , Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, We iss H, Xie Z. (2008). Single molecule DNA sequencing of a viral genome. *Science* **320**: 106-109.

Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Meijia JA, Booser D, Theriault RL, Buzdar AU, Dempsey PJ et al. (2006). Pharmacogenomic predictor of sensitivity to preoperative c hemotherapy with paclitaxel and fluorouracil, do xorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24**: 4236-4244.

Hook S.S., Lin J.J. and Dutta A. (2007). Mechanisms to control rereplication and implications for cancer. Curr. Opin. Cell Biol. 19: 663-671.

Hurtado A, Holmes KA, Ross-Innes CS, Schm idt D and Carroll JS (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nature Genetics 43: 27-33.

Karnani N., Taylor C., Malhotra A. and Du tta A. (2007). Pan-S replication patterns and chromosomal domains defin ed by ge nome-tiling arrays of E NCODE g enomic areas . Genome Res. 17: 685-676.

Lebofsky R. and W alter J.C. (2007). New Myc-anisms fo r DNA replication and tumorigenesis? Cancer Cell 12: 102-103.

Lin CY, Vega VB et al. (2007). Whole-genome catography of estrogen receptor alpha binding sites. PLoS Genet. 3(6): e87.

Lucas I., Palakodeti A., Jiang Y., Young D.J., Jiang N., Fernald A. A. and LeBeau M.M. (2007). High-throughput mapping of origins of replication in human cells. EMBO Rep. 8: 770-777.

Mesner LD, Valsakumar V, Karnani N, Dutta A, Hamlin J and Bekiranov S (2010) Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res. Dec 20 (Epub ahead of print).

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen A, Gerald WL and Massague J (2005). *Nature* **436**: 518-524.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A and Gray JW (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**: 515-527.

Nguyen V. Q., Co C. and Li J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature 411: 1068-1073.

Nikolsky Y., Sviridov E., Yao J., Dosymbekov D., Ustyansky V., Kaznacheev V., Dezso Z., Mulvey L., Macconaill L.E., Winckler W., Serebryiskaya T., Nikolskaya T. and Polyak K. (2008). Genome-wide functional synergy between amplified and mutated genes in human breast cancer. Cancer Res. 68: 9532-9540.

Olshen A. B., Venkatraman E.S., Lucito R. and Wigler M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5: 557-572.

Pan H., Deng, Y. and Pollard J.W. (2006). Progesterone blocks estrogen-induced DNA synthesis through the inhibition of replication licensing. Proc. Nat. Acad. Sci. 103: 14021-14026.

Petropoulou C., Kotantaki P., Karamitros D. and Tara Viras S. (2008). Cdt1 and Geminin in cancer: markers or triggers of malignant transformation? Front. Biosci. 13: 4485-4494.

Raphael B.R., Volik S., Yu P. et al. (2008). A sequence-based survey of the complex structural organization of tumor genomes. Genome Biol. 9: R59

Raveendranathan M., Chattopadhyay S., Bolon Y.T., Haworth J., Clarke D.J. and Bielinsky A.K. (2006). Genome-wide replication profiles of S-phase checkpoint mutants reveal fragile sites in yeast. EMBO J. 25: 3627-3639.

Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P et al. (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**: 166-180.

Shi Y-K., Yu Y.P., Zhu Z-H., Han Y-C., Rec B., Nelson J.B. and Luo J-H. (2008). MCM7 interacts with androgen receptor. Amer. J. Path. 173: 1758-1767.

Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL and Liu ET (2003). Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci* **100**: 10393-10398.

Teer J.K. and Dutta A. (2008). Human Cdt1 lacking the evolutionarily conserved region that interacts with MCM2-7 is capable of inducing re-replication. J. Biol. Chem. 283: 6817-6825.

Tao L., Dong Z., Leffak M., Zannis-Hadjopoulos M. and Price G. (2000). Major DNA replication initiation sites in the c-myc locus in human cells. J. Cell. Biochem. 78, 442-457.

Tomlins S.A., Rhodes D.R., Perner S. et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310: 644-648.

Volik S., Zhao S., Chin K. et al. (2003). End-sequence profiling: sequence-based analysis of aberrant genomes. Proc. Nat. Acad. Sci. 100: 7696-7701.

Volik S., Raphael B.J., Huang, G. et al. (2006). Decoding the fine-scale structure of a breast cancer genome and transcriptome. Genome Res.16: 394-404.

Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talentov D, Timmermans M, Meijer-van Gelder ME, Yu J et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**: 671-679.

Welboren WJ, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN and Stunnenberg HG (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. EMBO J. 28: 1418-1428.

Zhu W. and Dutta A. (2006). Activation of Fanconi anemia pathway in cells with re-replicated DNA. Cell Cycle 5: 2306-2309.

APPENDICES

Appendix 1: Meeting Abstracts

Meeting abstract: DOD 2008 Era of Hope meeting (Baltimore, MD)

Hormonal Involvement in Breast Cancer Gene Amplification

Michael S. Foulk[1*], Sara Hillenmeyer*[1], Alexander S. Brodsky[1], Benjamin J. Raphael[1], Shamlal Mangray[2], Theresa Graves[2] and <u>Susan A. Gerbi[1]</u>

[1] Brown University, Providence, RI 02912
[2] Lifespan (RI Hospital), Providence, RI 02903
* equal contribution

Genetic instability and rearrangements, including gene amplification, is a hallmark of cancer. Amplification of the HER2 (ErbB2/Neu) gene occurs in invasive breast cancer (~25%) and in ductal carcinoma *in situ* (50-60%). HER2 amplification and concomitant over-expression of this growth factor promotes cancer cell growth, acting as a metastasis-promoting factor. It would be desirable to prevent HER2 gene amplification, thereby moderating the aggressive growth of breast cancer cells. The problem is that no one knows what triggers gene amplification. Our recent research suggested that the trigger may be the steroid hormone estrogen. Do genetic or epigenetic changes produce novel binding site(s) for the estrogen receptor (ER) near the HER2 replication origin, inducing gene amplification? Our hypothesis is that ER interacts with the replication machinery to drive re-replication of the HER2 locus, resulting in DNA amplification.

Our specific aims and the study design are:

(1) Map ER binding sites in surgically derived HER2 amplified breast cancer tissue, using chromatin immunoprecipitation (ChIP) with an antibody against ER. The immunoprecipitated DNA will be used as a probe for DNA microarray chips ("ChIP-chip") to screen the human genome for hormone receptor binding sites. We will look for differences in ER binding sites between cancer cells and non-cancer cells from the same patient. The positive candidates will be confirmed by quantitative PCR following ChIP.

(2) Map replication origins using short nascent strands as probes for DNA microarray chips. Data analysis will identify replication origins that are near ER binding sites, with special attention given to novel ER sites in the cancer genome. An alternate and/or confirmatory approach is sequential ChIP ("re-ChIP") on chip experiments where DNA is immunoprecipitated by antibodies against ER and against Origin Recognition Complex polypeptide 2 (ORC2), thereby pulling down DNA fragments bound by both antigens.

(3) Quantify level of HER2 amplification and identify sites of co-amplification in the genome. DNA will be isolated from the same tissue samples used for specific aims (1) and (2) for use as probes for whole genome SNP arrays to quantify gene copy

numbers, thereby identifying regions of amplification. The level of HER2 gene amplification will be quantified and any sites of co-amplification will be determined. This study will examine if a correlation exists between ER binding at novel sites in the breast cancer genome and juxtaposition with putative replication origins that escape normal cellular controls and re-replicate, leading to DNA amplification. This may provide a new paradigm for hormonal induction of breast cancer via gene amplification, leading to new methods of diagnosis and treatment. Our results will indicate if there are other regions that co-amplify with the HER2 locus in the ER positive, HER2 amplified breast cancer patient samples. Other co-amplified genes, within the HER2 amplicon and/or at other regions, could serve as additional novel targets for therapies similar to the approach of using Herceptin to target HER2.

SUPPORTING DATA

Figures 1-6
Tables 1-2

Figures

Figure 1:

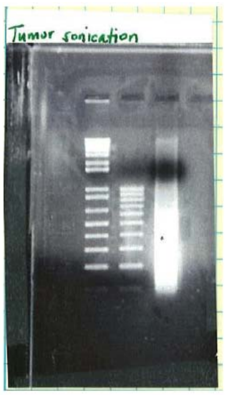

Figure 1: Representative gel showing good shearing of genomic DNA from an ER+ breast tumor sample using the Bioruptor. The largest signal observed is in the range of 500 bp.

Figure 2:

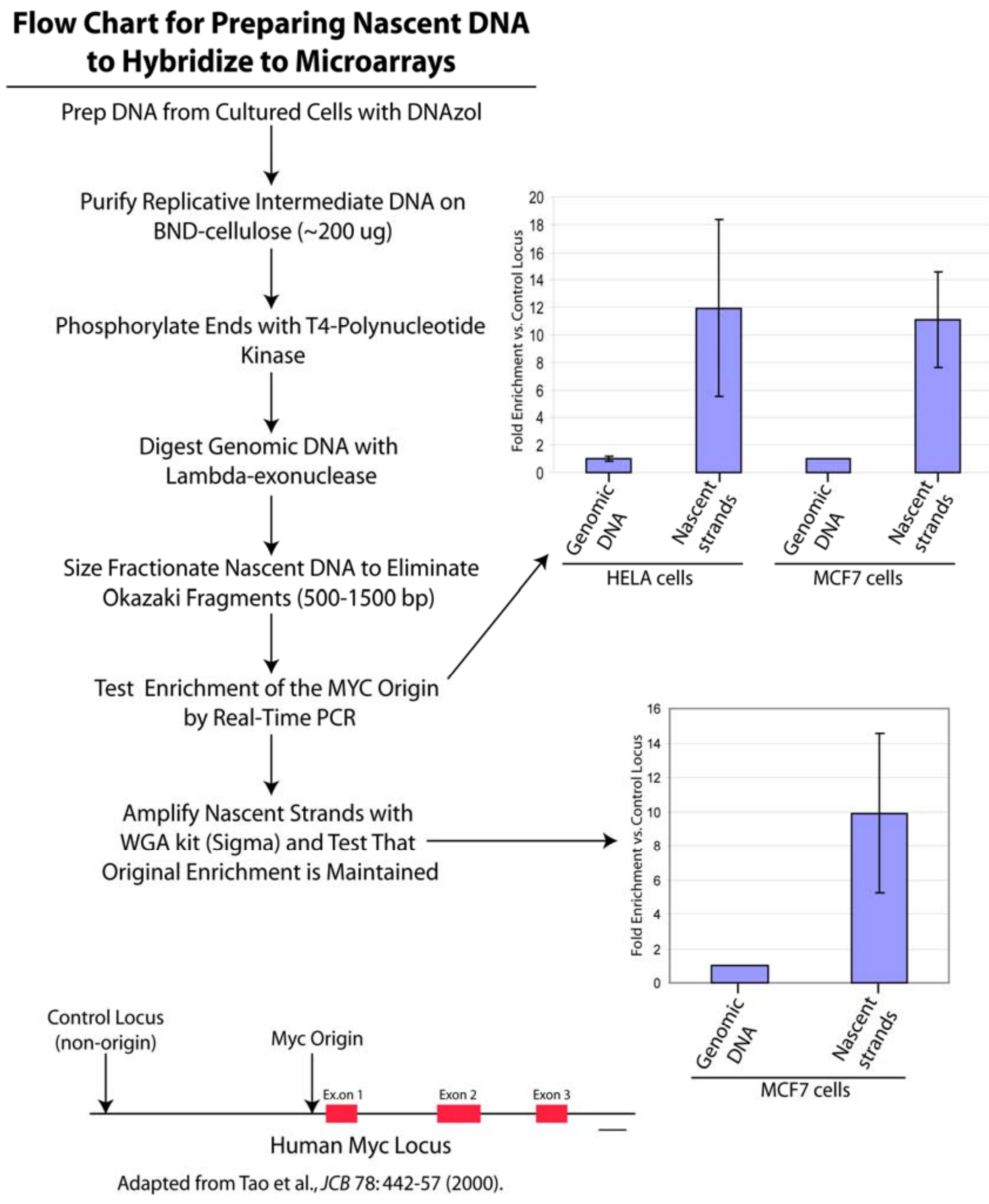

Figure 2: Flow chart of the nascent strand preparation protocol. At the bottom is a cartoon of the c-myc locus. The arrows indicate the targets of the real-time PCR primers. At right is real-time PCR data showing enrichment of the c-myc origin in both HeLa and MCF7 cells (top) and the maintenance of enrichment after WGA amplification of ht nascent strands (bottom).

Figure 3:

Figure 3: Chart of Illumina reads from MCF7 cell nascent strands mapped to the 283 origins identified by Cadoret et al., 2008. The red line indicates the cut off for statistical significance.

Figure 4:

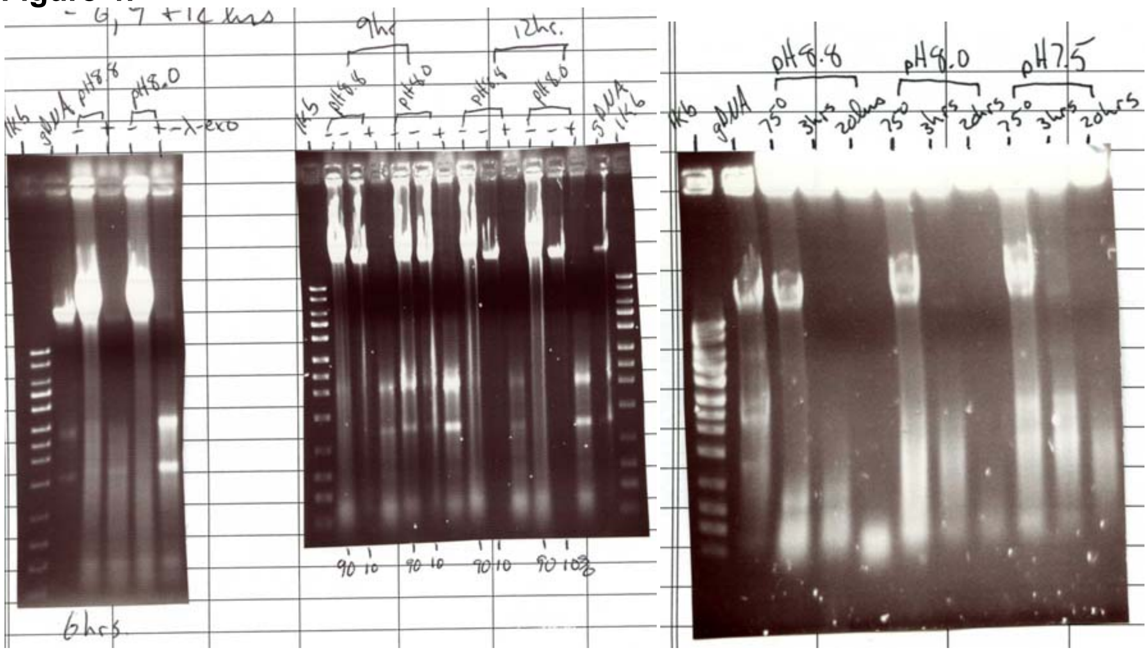

Figure 4: Gels from the notebook showing that the Fermentas lambda exonuclease is active at pH 8.0 and that the rRNA is more stable at this pH over the course of the digestion. 5 ug of DNA/RNA was digested at the indicated pH and for the time indicated. In the middle gel, the undigested sample was loaded in two adjacent lanes with one receiving 90% of the sample and the other the remaining 10%. The gel on the right shows that heating the sample to 75º C degrades the rRNA despite the pH (75º lanes).

Figure 5:

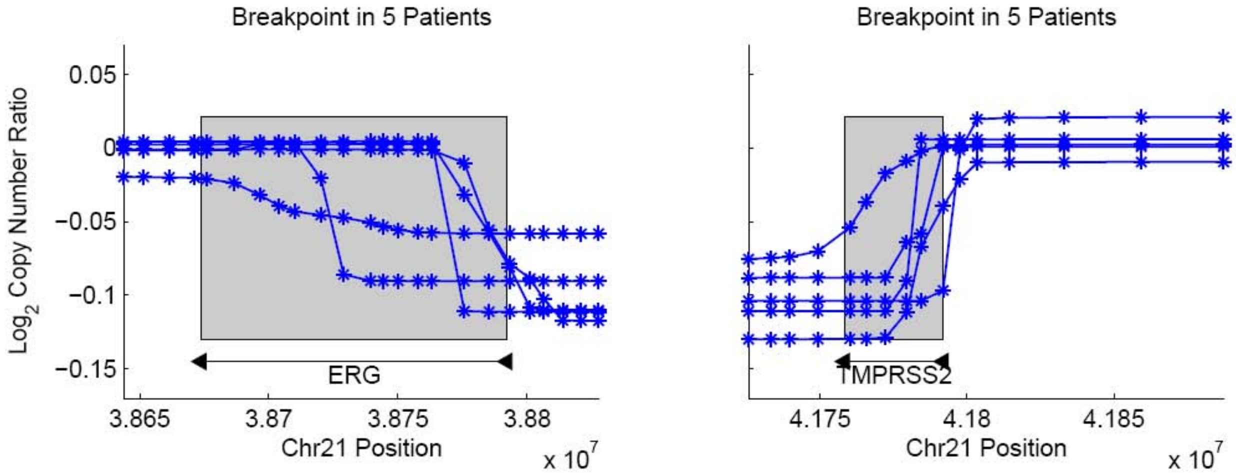

Figure 5: We identify the TMPRSS2-ERG fusion gene in 5 prostate cancer patients. The segmentations for each patient are shown in blue, and the asterisks denote probe locations. The deletion fuses the 5' end of TMPRSS2 to the 3' end of ERG, and the relative copy number at these breakpoints is conserved across the deletion.

Figure 6:

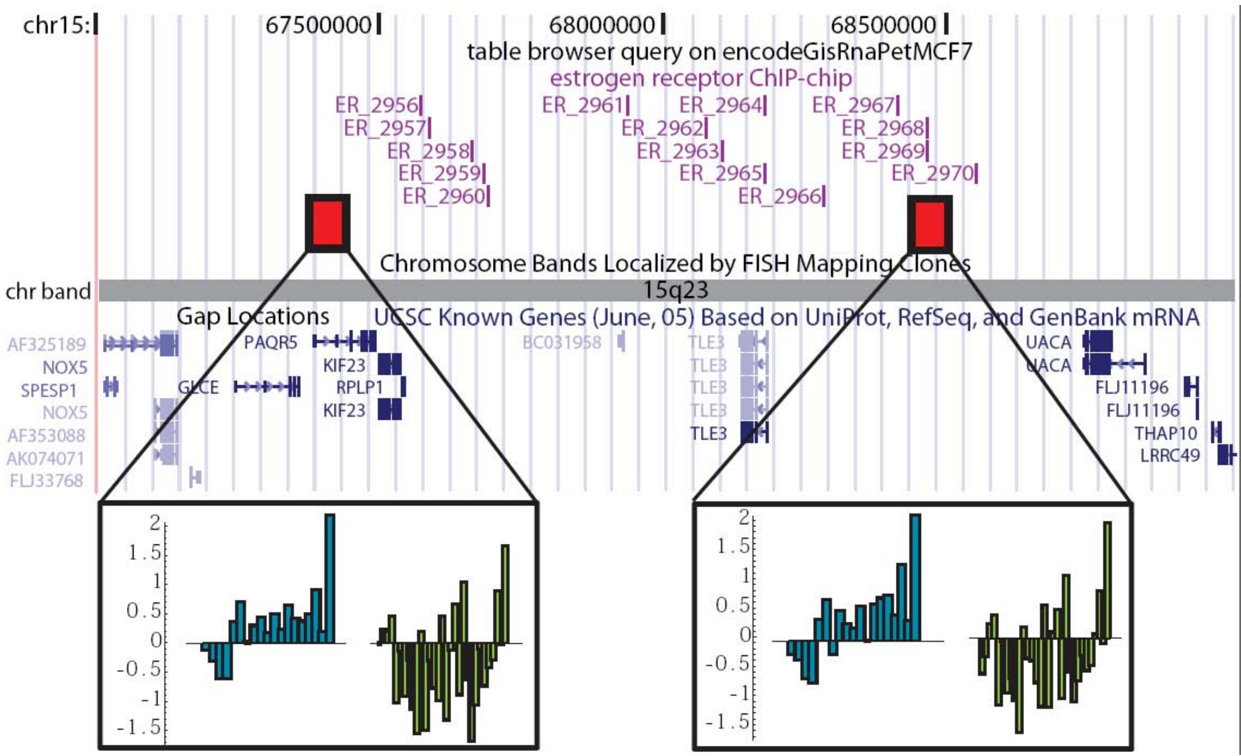

Figure 6: A statistica lly significant cluster of 15 ER bind ing sites (p urple lines, top) identified on Chromosome 15.

TABLE 1: Log of Breast Cancer Tumor Samples Received from R.I. Hospital

All tumor specimens were provided by Dr. Shamlal Mangray (Pathology Department, RI Hospital) and were frozen at -80 degrees. All samples were from female patients (identity unknown – coded by the Pathology Department) without neoadjuvent chemotherapy. Most were patients of Dr. Thersa Graves. The samples were 1.0-1.5 cm.

#	code	Date	ER	PR	HER2	Age	Comments
33	1	10/25/07	pos				used for H4 test1
34	2	11/13/07					also normal tissue
35	3	12/17/07	pos	pos	neg		also normal tissue
36	4 Path # 213A/J	1/18/08				post-menopausal(PM)	normal tissue (tube J)
37	5						
38	6 Path # 375C/F	1/18/08				45	used:C=H4 test (1 g) also normal tissue (tube F)
39	7						
40	8	1/28/08	pos	pos	neg	60(PM)	also normal tissue
41	9						
42	10 (SG5)	1/28/08	pos	pos	neg	49	used: H4/ER/mock(0.2 g) also normal tissue
43	11						
44	12 (SG6)	1/28/08	pos	pos	neg	41	also normal tissue
3	SG8	9/09	3+	2+	Neg	73	8 cm tumor, lymph node mets
5	SG10 (5-10%)	9/09	3+	1+	Neg	28	4 cm tumor, sentinel lymph node (SLN) micromets
6	SG11	9/09	3+	3+	Neg	54	2.5 cm tumor, no mets to SLN
8	SG13	9/09	Neg	Neg	Neg	53	1.7 cm tumor,

							no mets to SLN
9	SG14	9/09	3+	3+	2+(FISH neg)	33	4 cm tumor no mets to SLN
10	SG15	9/09	3+	3+	2+(FISH neg)	66	2.4 cm tumor, No SLN sampling, FNA of node negative
11	SG16	9/09	Neg	Neg	2+(FISH neg)	80	1.1 cm tumor, no mets to SLN
12	SG17	9/09	3+	Neg	2+(FISH neg)	69	2.5 cm tumor, Axillary lymph node (ALN) negative
13	SG18	9/09	3+	3+	FISH neg	59	1.3 cm tumor, no mets to SLN
4	19	11/8/10	pos	pos	neg	also	normal tissue
5	20	11/8/10	pos	pos	neg	also	normal tissue
6	21	11/8/10	pos	pos	neg	also	normal tissue
7	22	11/8/10	pos	pos	neg	also	normal tissue
8	25	11/8/10	pos	pos	neg	also	normal tissue
9	26	11/8/10	pos	pos	neg	also	normal tissue
10	27	11/8/10	pos	pos	neg	also	normal tissue
11	28	11/8/10	pos	pos	neg	also	normal tissue
12	29	11/8/10	pos	pos	neg	also	normal tissue
13	30	11/8/10	pos	pos	neg	also	normal tissue
14	31	11/8/10	pos	pos	neg	also	normal tissue

TABLE 2: Spreadsheet of MCF7 nasc ent DNA Illumina re ads mapped to the ENCODE data set.

	A	B	C	D	E	F	G	H	I
1	Origi n - Chro m	Origin Start	Origin End	Num Rea ds	Reads/ kb	pval (Poisson)	Corrected pval	Mapp ed	Mapped/ kb
2	chr1 1	6413111 0	6413298 7 37		19.712 31 0		0	26510 49 0.94	680
3	chr1 1	6428967 4	6429300 5 75		22.515 76	0	0		
4	chr1 1	6432689 7	6432868 9 35		19.531 25	0	0		
5	chr1 1	6436748 6	6436864 8 34		29.259 90	0	0		
6	chr1 3	1125965 65	1125979 31 27		19.765 74	0	0		
7	chr1 6	154879 1	56172 22		17.014 69	0	0		
8	chr1 6	271684 2	72766 25		23.105 36	0	0		
9	chr1 6	343083 3	44285 27		22.462 56	0	0		
10	chr1 6	353862 3	55393 26		16.982 36	0	0		
11	chr2	2200577 05	2200591 50 27		18.685 12	0	0		
12	chr2	2200812 16	2200833 96 37		16.972 48	0	0		
13	chr2	2200846 25	2200858 41 29		23.848 68	0	0		
14	chr2	2201111 58	2201130 02 39		21.149 67	0	0		
15	chr2	2201247 53	2201282 35 81		23.262 49	0	0		
16	chr2	2201332 14	2201346 42 25		17.507 00	0	0		
17	chr2 1	3960765 5	3960878 9 24		21.164 02	0	0		
18	chr5	1316209 57	1316223 17 53		38.970 59	0	0		
19	chr6	4162251	4162406	46	29.715	0	0		

		3 1			76				
20	chr9	130813941	130816102 37		17.12170	0	0		
21	chr9	130830374	130831123 19		25.36716	0	0		
22	chrX	152889595	152890789 38		31.82580	0	0		
23	chrX	153339431	153340844 33		23.35456	0	0		
24	chr16	265752	267214 23		15.73187	0.0000000000001	0.00000000004		
25	chr2	219989990	219991533 24		15.55412	0.0000000000001	0.00000000004		
26	chr2	220048583	220051656 47		15.29450	0.0000000000001	0.00000000004		
27	chr20	33652973	33654116 18		15.74803	0.0000000000001	0.00000000004		
28	chr6	108385466	108386797 20		15.02630	0.0000000000001	0.00000000004		
29	chr7	27119915	27121439 24		15.74803	0.0000000000001	0.00000000004		
30	chr9	130890421	130891863 22		15.25659	0.0000000000001	0.00000000004		
31	chr9	130910497	130915036 70		15.42190	0.0000000000001	0.00000000004		
32	chrX	152853510	152854717 19		15.74151	0.0000000000001	0.00000000004		
33	chr11	64167296	64168971 24		14.32836	0.000000000002	0.0000000006		
34	chr14	98767316	98769158 27		14.65798	0.000000000002	0.0000000006		
35	chr16	47415	48530 16		14.34978	0.000000000002	0.0000000006		
36	chr19	59174913	59178718 55		14.45466	0.000000000002	0.0000000006		
37	chr9	130980491	130981677 17		14.33390	0.000000000002	0.0000000006		
38	chr14	98808316	98809543 17		13.85493	0.00000000003	0.000000009		
39	chr16	371926	373475 21		13.55713	0.00000000003	0.000000009		

40	chr2	220018396	220020076 22		13.09524	0.00000000003	0.000000009		
41	chr5	131860038	131861573 20		13.02932	0.00000000003	0.000000009		
42	chr11	64154782	64156726 24		12.34568	0.0000000005	0.0000001		
43	chr14	98782269	98783573 16		12.26994	0.0000000005	0.0000001		
44	chr18	24011725	24012891 14		12.00686	0.0000000005	0.0000001		
45	chr19	59062973	59064186 15		12.36603	0.0000000005	0.0000001		
46	chr6	41854596	41855697 14		12.71571	0.0000000005	0.0000001		
47	chr7	27114055	27115979 24		12.47401	0.0000000005	0.0000001		
48	chr7	27135068	27136514 18		12.44813	0.0000000005	0.0000001		
49	chr9	131180525	131181712 15		12.63690	0.0000000005	0.0000001		
50	chr11	64302035	64303808 20		11.28032	0.000000006	0.000002		
51	chr7	27225347	27227050 20		11.74398	0.000000006	0.000002		
52	chr7	90063542	90064916 16		11.64483	0.000000006	0.000002		
53	chr9	131138758	131140299 18		11.68073	0.000000006	0.000002		
54	chrX	153326081	153327218 13		11.43360	0.000000006	0.000002		
55	chrX	153416720	153418174 16		11.00413	0.000000006	0.000002		
56	chr11	64249506	64251467 20		10.19888	0.00000007	0.00002		
57	chr20	33667226	33668962 19		10.94470	0.00000007	0.00002		
58	chr7	27160404	27161510 12		10.84991	0.00000007	0.00002		
59	chrX	153428085	153429458 14		10.19665	0.00000007	0.00002		
60	chr1	149697853	149699882 19		9.36422	0.0000007	0.00020		

61	chr11	2148231	2149404	11	9.37766	0.0000007	0.00020		
62	chr11	64438527	64439948 14		9.85222	0.0000007	0.00020		
63	chr21	39299170	39300343 11		9.37766	0.0000007	0.00020		
64	chr6	41578447	41579719 12		9.43396	0.0000007	0.00020		
65	chrX	152822036	152824175 21		9.81767	0.0000007	0.00020		
66	chr1	149750298	149751652 12		8.86263	0.000007	0.00196		
67	chr11	64160072	64161187 10		8.96861	0.000007	0.00196		
68	chr19	59385445	59387234 16		8.94354	0.000007	0.00196		
69	chr2	220144495	220145916 12		8.44476	0.000007	0.00196		
70	chr21	33314649	33316910 19		8.40336	0.000007	0.00196		
71	chr5	132025755	132026955 10		8.33333	0.000007	0.00196		
72	chr6	41518826	41520409 14		8.84397	0.000007	0.00196		
73	chr6	41856075	41857739 14		8.41346	0.000007	0.00196		
74	chr7	27116146	27118019 15		8.00854	0.000007	0.00196		
75	chr7	27240766	27242195 12		8.39748	0.000007	0.00196		
76	chr9	130838587	130840206 13		8.02965	0.000007	0.00196		
77	chrX	152948833	152950748 17		8.87728	0.000007	0.00196		
78	chrX	153016122	153017675 13		8.37090	0.000007	0.00196		
79	chrX	153256057	153257154 9		8.20419	0.000007	0.00196		
80	chr11	64368631	64369884 10		7.98085	0.00006	0.01676		
81	chr11	64412337	64413999 13		7.82190	0.00006	0.01676		

82	chr1 3	1127529 96	1127542 03 9		7.4565 0	0.00006	0.01676		
83	chr1 4	9848627 8	9848775 7 11		7.4374 6	0.00006	0.01676		
84	chr1 6	176977 1	77951 7		7.1868 6	0.00006	0.01676		
85	chr1 6	391235 3	92389 9		7.7989 6	0.00006	0.01676		
86	chr2 2	3035513 5	3035625 5 8		7.1428 6	0.00006	0.01676		
87	chr6	4165412 0	4165575 0 12		7.3619 6	0.00006	0.01676		
88	chr7	1159514 28	1159529 38 12		7.9470 2	0.00006	0.01676		
89	chr9	1309425 15	1309456 15 22		7.0967 7	0.00006	0.01676		
90	chr9	1312141 69	1312164 97 17		7.3024 1	0.00006	0.01676		
91	chr1 3	1128112 87	1128125 98 9		6.8649 9	0.00045	0.12623		
92	chr1 4	9888185 9	9888320 2 9		6.7014 1	0.00045	0.12623		
93	chr1 6	144166 1	45603 9		6.2630 5	0.00045	0.12623		
94	chr1 9	5909347 0	5909564 7 15		6.8902 2	0.00045	0.12623		
95	chr1 9	5918659 0	5918779 0 8		6.6666 7	0.00045	0.12623		
96	chr1 9	5963361 4	5963465 6 7		6.7178 5	0.00045	0.12623		
97	chr2 1	3332165 8	3332345 7 11		6.1145 1	0.00045	0.12623		
98	chr2 1	3344575 2	3344685 9 7		6.3234 0	0.00045	0.12623		
99	chr2 2	3084905 1	3085026 5 8		6.5897 9	0.00045	0.12623		
10 0	chr2 2	3090151 6	3090271 5 8		6.6722 3	0.00045	0.12623		
10 1	chr6	7416071 3	7416199 2 8		6.2548 9	0.00045	0.12623		
10 2	chr6	1085474 74	1085487 73 9		6.9284 1	0.00045	0.12623		

10 3	chr7	2717609 7	2717727 2 8		6.8085 1	0.00045	0.12623		
10 4	chr9	1310860 93	1310880 57 12		6.1099 8	0.00045	0.12623		
10 5	chrX	1527943 89	1527956 60 8		6.2942 6	0.00045	0.12623		
10 6	chr1	1497120 58	1497137 85 9		5.2113 5	0.00291	0.81993		
10 7	chr1 1	2127225	2128288	6	5.6444 0	0.00291	0.81993		
10 8	chr1 3	1124695 91	1124711 44 8		5.1513 2	0.00291	0.81993		
10 9	chr1 3	1127039 46	1127058 45 11		5.7925 2	0.00291	0.81993		
11 0	chr1 4	9872431 4	9872564 5 7		5.2592 0	0.00291	0.81993		
11 1	chr1 9	5920695 6	5920815 6	6	5	0.00291	0.81993		
11 2	chr5	5624119 9	5624262 5 8		5.6101 0	0.00291	0.81993		
11 3	chr7	2725001 4	2725117 6 6		5.1635 1	0.00291	0.81993		
11 4	chr7	1145446 82	1145457 45 6		5.6444 0	0.00291	0.81993		
11 5	chr9	1311535 33	1311559 49 14		5.7947 0	0.00291	0.81993		
11 6	chrX	1228217 73	1228228 61 6		5.5147 1	0.00291	0.81993		
11 7	chrX	1532137 03	1532149 59 7		5.5732 5	0.00291	0.81993		
11 8	chr1 1	1309325 56	1309336 63 5		4.5167 1	0.01590	4.48334		
11 9	chr1 1	1310477 67	1310492 30 7		4.7846 9	0.01590	4.48334		
12 0	chr1 3	1124282 05	1124296 22 6		4.2343 0	0.01590	4.48334		
12 1	chr1 3	1125301 34	1125318 63 8		4.6269 5	0.01590	4.48334		
12 2	chr1 3	1126459 63	1126471 84 6		4.9140 0	0.01590	4.48334		
12 3	chr1 3	1126689 83	1126706 26 8		4.8691 4	0.01590	4.48334		

124	chr19	59138068	59139337 6		4.72813	0.01590	4.48334		
125	chr19	59357776	59359107 6		4.50789	0.01590	4.48334		
126	chr2	220424166	220425867 8		4.70312	0.01590	4.48334		
127	chr2	234369318	234370687 6		4.38276	0.01590	4.48334		
128	chr21	33697620	33699322 8		4.70035	0.01590	4.48334		
129	chr22	30807961	30809710 7		4.00229	0.01590	4.48334		
130	chr5	131636239	131637448 5		4.13565	0.01590	4.48334		
131	chr5	142078249	142081176 14		4.78305	0.01590	4.48334		
132	chr6	41599804	41601172 6		4.38596	0.01590	4.48334		
133	chr7	89712535	89713744 6		4.96278	0.01590	4.48334		
134	chr7	115926787	115928780 9		4.51581	0.01590	4.48334		
135	chr7	115953373	115954646 6		4.71328	0.01590	4.48334		
136	chr7	116750268	116751510 6		4.83092	0.01590	4.48334		
137	chr7	126774419	126775775 6		4.42478	0.01590	4.48334		
138	chr7	126819852	126821344 7		4.69169	0.01590	4.48334		
139	chrX	152836589	152837735 5		4.36300	0.01590	4.48334		
140	chrX	153202702	153203848 5		4.36300	0.01590	4.48334		
141	chr1	149813684	149814915 4		3.24939	0.07078	19.96030		
142	chr13	112751295	112752716 5		3.51865	0.07078	19.96030		
143	chr19	59104275	59106633 8		3.39271	0.07078	19.96030		
144	chr21	33773906	33775426 6		3.94737	0.07078	19.96030		

145	chr21	33937141	33938736 5		3.13480	0.07078	19.96030		
146	chr22	30341875	30342832 3		3.13480	0.07078	19.96030		
147	chr22	31200775	31202817 8		3.91773	0.07078	19.96030		
148	chr5	56147525	56149060 5		3.25733	0.07078	19.96030		
149	chr5	142176213	142178174 7		3.56961	0.07078	19.96030		
150	chr6	41543060	41544342 5		3.90016	0.07078	19.96030		
151	chr6	41544942	41546022 4		3.70370	0.07078	19.96030		
152	chr6	41585098	41586734 6		3.66748	0.07078	19.96030		
153	chr6	41719178	41720147 3		3.09598	0.07078	19.96030		
154	chr6	74287525	74288636 4		3.60036	0.07078	19.96030		
155	chr7	27222016	27223315 4		3.07929	0.07078	19.96030		
156	chr7	116290006	116291315 5		3.81971	0.07078	19.96030		
157	chr7	116381043	116382402 5		3.67918	0.07078	19.96030		
158	chr7	127016091	127017258 4		3.42759	0.07078	19.96030		
159	chr9	130982808	130984393 5		3.15457	0.07078	19.96030		
160	chr13	112395220	112397014 4		2.22965	0.24468	68.99993		
161	chr13	112561443	112562786 3		2.23380	0.24468	68.99993		
162	chr13	112606307	112607524 3		2.46508	0.24468	68.99993		
163	chr13	112684521	112686329 4		2.21239	0.24468	68.99993		
164	chr14	98696258	98697981 5		2.90192	0.24468	68.99993		
165	chr19	59077897	59080362 6		2.43408	0.24468	68.99993		

16 6	chr1 9	5934000 3	5934136 8 3		2.1978 0	0.24468	68.99993		
16 7	chr2	2200715 68	2200741 68 7		2.6923 1	0.24468	68.99993		
16 8	chr2	2345541 79	2345555 16 3		2.2438 3	0.24468	68.99993		
16 9	chr2 0	3360464 7	3360627 9 4		2.4509 8	0.24468	68.99993		
17 0	chr2 1	3384729 4	3384917 3 4		2.1287 9	0.24468	68.99993		
17 1	chr5	5608806 6	5608947 8 3		2.1246 5	0.24468	68.99993		
17 2	chr5	1315718 51	1315736 44 5		2.7886 2	0.24468	68.99993		
17 3	chr5	1318011 34	1318028 91 5		2.8457 6	0.24468	68.99993		
17 4	chr5	1320481 84	1320494 26 3		2.4154 6	0.24468	68.99993		
17 5	chr5	1422362 58	1422372 49 2		2.0181 6	0.24468	68.99993		
17 6	chr7	2713713 6	2713833 0 3		2.5125 6	0.24468	68.99993		
17 7	chr8	1192501 80	1192514 04 3		2.4509 8	0.24468	68.99993		
17 8	chrX	1227262 13	1227275 33 3		2.2727 3	0.24468	68.99993		
17 9	chrX	1229231 20	1229244 70 4		2.9629 6	0.24468	68.99993		
18 0	chrX	1532624 07	1532638 80 3		2.0366 6	0.24468	68.99993		
18 1	chr1 3	2989390 5	2989541 3 3		1.9893 9	0.61202	172.58983		
18 2	chr1 4	5298863 3	5299017 1 2		1.3003 9	0.61202	172.58983		
18 3	chr1 6	128359 1	29221 1		1.1600 9	0.61202	172.58983		
18 4	chr1 6	458042 4	59289 2		1.6038 5	0.61202	172.58983		
18 5	chr1 9	5906535 8	5906716 9 3		1.6565 4	0.61202	172.58983		
18 6	chr1 9	5915533 0	5915717 3 3		1.6277 8	0.61202	172.58983		

187	chr20	33520581	33521619 2		1.92678	0.61202	172.58983		
188	chr20	33593121	33594722 3		1.87383	0.61202	172.58983		
189	chr21	32736476	32738137 2		1.20409	0.61202	172.58983		
190	chr21	33143912	33145612 3		1.76471	0.61202	172.58983		
191	chr21	39377239	39378501 2		1.58479	0.61202	172.58983		
192	chr22	31035855	31036860 2		1.99005	0.61202	172.58983		
193	chr22	31371562	31373126 2		1.27877	0.61202	172.58983		
194	chr5	55873281	55874378 2		1.82315	0.61202	172.58983		
195	chr5	141999954	142001505 2		1.28949	0.61202	172.58983		
196	chr5	142361833	142363780 2		1.02722	0.61202	172.58983		
197	chr6	74108528	74110397 3		1.60514	0.61202	172.58983		
198	chr6	108711089	108712249 2		1.72414	0.61202	172.58983		
199	chr7	89870825	89872426 2		1.24922	0.61202	172.58983		
200	chr7	116591693	116593404 2		1.16891	0.61202	172.58983		
201	chr7	116906722	116908786 4		1.93798	0.61202	172.58983		
202	chr7	117218481	117219654 2		1.70503	0.61202	172.58983		
203	chr8	119302424	119303820 2		1.43266	0.61202	172.58983		
204	chr9	130824308	130825849 2		1.29786	0.61202	172.58983		
205	chr9	131164410	131165386 1		1.02459	0.61202	172.58983		
206	chrX	122763000	122764146 2		1.74520	0.61202	172.58983		
207	chr1	149438460	149440397	0	0	1	282		

208	chr11	1715184	1716695	1	0.66181	1	282		
209	chr11	130855483	130856606 1		0.89047	1	282		
210	chr12	38937995	38939318 1		0.75586	1	282		
211	chr13	29850448	29851781 1		0.75019	1	282		
212	chr13	112433768	112434909 1		0.87642	1	282		
213	chr13	112602511	112604575 2		0.96899	1	282		
214	chr13	112612988	112614632 1		0.60827	1	282		
215	chr14	98771506	98772663 1		0.86430	1	282		
216	chr14	98956703	98958083 1		0.72464	1	282		
217	chr15	41735285	41736157	0	0	1	282		
218	chr15	41850700	41852158	0	0	1	282		
219	chr15	42013413	42014920 1		0.66357	1	282		
220	chr16	60988052	60989443	0	0	1	282		
221	chr18	24064501	24065770	0	0	1	282		
222	chr19	59225201	59226338 1		0.87951	1	282		
223	chr19	59588883	59590167 1		0.77882	1	282		
224	chr2	51885784	51887168 1		0.72254	1	282		
225	chr2	118506164	118507311	0	0	1	282		
226	chr2	220218828	220219997 1		0.85543	1	282		
227	chr2	220221572	220222989 1		0.70572	1	282		
228	chr2	234539246	234540367 1		0.89206	1	282		

229	chr2	234549286	234550644	0	0	1	282		
230	chr2	234623194	234624340	0	0	1	282		
231	chr20	33449379	33451062 1		0.59418	1	282		
232	chr20	33763526	33764778 1		0.79872	1	282		
233	chr21	32825833	32827210 1		0.72622	1	282		
234	chr21	32924855	32926140	0	0	1	282		
235	chr21	33371196	33372523 1		0.75358	1	282		
236	chr21	33869514	33870830 1		0.75988	1	282		
237	chr22	30319739	30320501	0	0	1	282		
238	chr22	31162231	31163655	0	0	1	282		
239	chr22	31442946	31444794	0	0	1	282		
240	chr22	31536159	31537121	0	0	1	282		
241	chr22	31550451	31551647 1		0.83612	1	282		
242	chr5	55981348	55982831 1		0.67431	1	282		
243	chr5	56085418	56086517	0	0	1	282		
244	chr5	56119168	56120680	0	0	1	282		
245	chr5	131744261	131745522 1		0.79302	1	282		
246	chr5	132151529	132152659 1		0.88496	1	282		
247	chr5	132248525	132250628 1		0.47551	1	282		
248	chr5	141950090	141951462	0	0	1	282		
249	chr5	142199652	142201376 1		0.58005	1	282		

250	chr5	142259552	142261293	0	0	1	282		
251	chr6	74011080	74012121	0	0	1	282		
252	chr6	108552017	108553295	0	0	1	282		
253	chr6	132399844	132401367 1		0.65660	1	282		
254	chr7	27145830	27147560 1		0.57803	1	282		
255	chr7	89671933	89673120 1		0.84246	1	282		
256	chr7	89716391	89718325	0	0	1	282		
257	chr7	90230960	90232415	0	0	1	282		
258	chr7	90283033	90284598 1		0.63898	1	282		
259	chr7	90434126	90435342 1		0.82237	1	282		
260	chr7	114067204	114068337	0	0	1	282		
261	chr7	114309758	114311461 1		0.58720	1	282		
262	chr7	114511851	114512914 1		0.94073	1	282		
263	chr7	115921833	115923372	0	0	1	282		
264	chr7	115932826	115934434	0	0	1	282		
265	chr7	115985825	115987901 2		0.96339	1	282		
266	chr7	116126066	116128112 2		0.97752	1	282		
267	chr7	116223147	116226022	0	0	1	282		
268	chr7	116494825	116495804	0	0	1	282		
269	chr7	116606707	116607969	0	0	1	282		
270	chr7	116618519	116619873	0	0	1	282		

271	chr7	116914162	116915647 1		0.67340	1	282		
272	chr7	116992550	116993726	0	0	1	282		
273	chr7	126272923	126274135	0	0	1	282		
274	chr7	126306177	126306834	0	0	1	282		
275	chr7	126440757	126442109	0	0	1	282		
276	chr7	126892241	126893357	0	0	1	282		
277	chr8	119012420	119013804	0	0	1	282		
278	chr8	119374375	119375434 1		0.94429	1	282		
279	chr9	131000600	131001795 1		0.83682	1	282		
280	chrX	122662145	122663518	0	0	1	282		
281	chrX	122693012	122694269 1		0.79554	1	282		
282	chrX	123061667	123064036	0	0	1	282		
283	chrX	153908351	153909390	0	0	1	282		